



Contribution ID: 92

Type: Oral

MPI support and improvements in gLite/EGEE - a report from the MPI Task Force

Thursday 15 April 2010 09:00 (30 minutes)

The MPI Task Force (MPI-TF) was established in response to calls from the EGEE user communities and site administrators during the MPI session at EGEE09. Improvements in the overall deployment and support of MPI on the EGEE infrastructure were demanded. The objectives of the MPI-TF are to provide solutions which will improve the testing, installation and configuration of MPI-enabled middleware, and to support greater MPI application usage on the Grid. We report on the work performed by the MPI-TF and report on the changes and improvements in MPI support delivered since its establishment.

Detailed analysis

The reported high failure rates for MPI jobs on the EGEE grid is a source of frustration for many of its users. At the EGEE09 conference, the Earth Science community reported that only 7 sites from 26 supporting both MPI and its virtual organisation ran their applications without errors. Computational Chemistry reported that only 53% of jobs ran successfully at the CompChem enabled sites when requiring 8 processors, and this dropped to 21% when requesting 16 to 64 processors. Similar issues were reported by the Astronomy and Astrophysics community.

The site administrators reported on problems associated with installing and configuring MPI at their institutes. This was backed up with a report from the MPI Working Group (MPI-WG), who gathered extensive feedback from a large sample of users and site administrators.

The provision of a single MPI solution for gLite is non-trivial. This is in part due to lack of good solid support from upstream OS providers, and in part due to the diverse site system setups, such as batch systems, multi-core worker nodes, and shared versus non-shared filesystems. We present a pragmatic approach to systems support.

Conclusions and Future Work

The MPI-WG delivered a set of recommendations to improve MPI support on EGEE based on user and site administrator feedback. The MPI-TF was established to deliver many of the short term improvements which could be dealt with within the lifetime of EGEE-III. However, some issues cannot be easily concluded within this timeframe. Firstly, full interoperability with other grid infrastructures and middlewares such as ARC and UNICORE. Secondly, the emergence of multi-core and hybrid systems needs to be evaluated - this requires a generic "Parallel" job type. These will be examined in the EGI era.

Impact

In December 2009, over 100 sites were publishing support for MPI on the EGEE infrastructure. Some 50 of these sites pass the MPI validation tests; these probe the range of basic MPI distributions available at each site. Most sites successfully support at least one version of MPI correctly. The introduction of critical SAM validation tests is expected to help isolate problems at the sites and contribute to the development of a "troubleshooting" guide. Updated user documentation covering numerous use cases and recipes will help users and developers avoid unforeseen issues and in turn help get the best out of the infrastructure. Similarly, much

needed updates to the MPI/gLite related middleware will accommodate the diverse systems setups currently support for non-MPI job execution.

The MPI-TF consists of representatives from multiple EGEE-III activities, including JRA1, NA4, SA1 and SA3, with oversight from the EGEE TMB. We shall provide an evaluation of the impact of the MPI-TF by measuring improvements to the number of sites actively supporting MPI, sites successfully passing MPI SAM tests, and by seeking feedback from the user community through their virtual organisations.

Keywords

MPI, User Support

URL for further information

<https://twiki.cern.ch/twiki/bin/view/EGEE/MpiTools>

Authors: CAMPOS PLASENCIA, Isabel (IFCA); WALSH, John (Trinity College Dublin/Grid-Ireland)

Co-authors: COGHLAN, Brian (Trinity College Dublin); VAN DOK, Dennis (NIKHEF); FERNÁNDEZ DEL CASTILLO, Enol (IFCA); ENGELBERTS, Jeroen (SARA)

Presenter: WALSH, John (Trinity College Dublin/Grid-Ireland)

Session Classification: Parallel Processing

Track Classification: Software services exploiting and/or extending grid middleware (gLite, ARC, UNICORE etc)